DOCUMENT RESUME

ED 315 445 TM 014 471

TITLE Maryland Functional Mathematics Test--II: Fall 1988 Administration. Report M-88. Technical Report.
INSTITUTION Maryland State Dept. of Education, Baltimore. Div. of Instruction.
PUB DATE Jun 89
NOTE 24p.
PUB TYPE Statistical Data (110) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Grade 9; *Graduation Requirements; High Schools; *High School Students; Mathematics Achievement; Mathematics Skills; *Mathematics Tests; Public Schools; *State Programs; Testing Programs; *Test Results
IDENTIFIERS *Maryland Functional Mathematics Test

ABSTRACT

The Maryland Functional Mathematics Test, Level II (MFMT-II) assesses Maryland students' mastery of functional mathematics skills that are described in the Declared Competencies Index published in 1980 by the Maryland State Department of Education (MSDE). The test is also used to certify students for graduation from high school and to identify students in need of assistance in improving their mathematics skills. Forms 270 through 279 of the 1988-89 edition of the MFMT-II were administered to 47,024 public school students in grades 9 through 12 during the period October 17 to November 4, 1988. Students in all 24 school districts in Maryland took the test for the first time in grade 9; those who took it in grades 10 through 12 were transfers or had previously failed the test. Separate sections of this paper describe and report on the development procedures, technical characteristics, and student performance results of the 1988 administration of the MFMT-II. More detailed technical information and information on test reliability and validity is available from the MSDE. The percentage of students who passed on the first try decreased slightly from 1987. However, the mean scale score of 352.9 for 1988 represents only a 0.1 increase over the mean scale score of 352.8 for 1987. Nine tables present test results and passing percentages by grade. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED315445

TECHNICAL REPORT

MARYLAND FUNCTIONAL MATHEMATICS TEST-II
FALL 1988 ADMINISTRATION
REPORT M-88
JUNE 1989

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

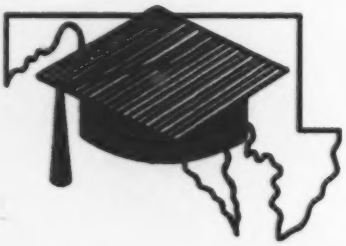
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. A. Crenson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."



Maryland State Department of Education
Division of Instruction
Program Assessment, Evaluation, and Instructional Support Branch
200 W. Baltimore Street
Baltimore, Maryland 21201
(301) 333-2369

MFTP

MARYLAND FUNCTIONAL TESTING PROGRAM

A COMPONENT OF PROJECT BASIC FOR:

* **ASSESSMENT OF BASIC AND CITIZENSHIP SKILLS**
* **IDENTIFICATION FOR APPROPRIATE ASSISTANCE**
* **CERTIFICATION FOR GRADUATION**

TM014471

BEST COPY AVAILABLE

STATE BOARD OF EDUCATION

Lawrence A. Shulman President	Rockville	1989
Wilson H. Parran Vice President	Huntingtown	1991
May B. Bolt	Frostburg	1989
Herbert Fincher	Salisbury	1993
Priscilla Hilliard	Upper Marlboro	1990
Rosetta G. Kerr	Baltimore	1991
Joan Clark Maynard	Linthicum Heights	1992
Frederick K. Schoenbrodt	Ellicott City	1990
Benjamin Swinson	Silver Spring	1992
Shirley Leung Student Member	Bel Air	1989

MARYLAND STATE DEPARTMENT OF EDUCATION

Joseph L. Shilling
State Superintendent of Schools

Claud E. Kitchens
Deputy State Superintendent of Schools

Martha J. Fields
Assistant Deputy State Superintendent

Nicholas Hobar
Assistant State Superintendent, Division of Instruction

James W. Latham
Director, Division of Instruction

Robert E. Gabrys
Chief, Program Assessment, Evaluation, and Instructional Support Branch

Steven Ferrara
Chief, Measurement, Statistics, and Evaluation Section

Leslie A. Walker-Bartnick
Specialist in Research and Evaluation

STATE OF MARYLAND

William Donald Schaefer
Governor

The Maryland State Department of Education does not discriminate on the basis of race, color, sex, age, national origin, religion, or handicapping conditions in matters affecting employment or in providing access to programs. For inquiries related to Departmental policy, contact the Office of Equal Opportunity.

MSDE
June 1989

Technical Report

MARYLAND FUNCTIONAL MATHEMATICS TEST-II
Fall 1988 Administration

Maryland State Department of Education
Division of Instruction
Program Assessment, Evaluation, and Instructional Support Branch

June 1989

The purposes of the Maryland Functional Mathematics Test, Level II (MFMT-II) are threefold. First, the test has been designed to assess Maryland students' mastery of functional mathematics skills that are described in the Declared Competencies Index (DCI) published by the Maryland State Department of Education (MSDE) in 1980. In addition, the test is used to certify students for high school graduation. Finally, the test is used to identify students who should participate in appropriate assistance in order to improve their mathematics skills.

Forms 270-279 of the 1988-89 edition of the Maryland Functional Mathematics Test, Level II were administered in Maryland public schools to students in grades 9-12 during the period October 17-November 4, 1988. Students in all of the state's 24 school districts took the MFMT-II for the first time in ninth grade. Grade 10, 11, and 12 students who took the 1988 MFMT-II include members of the graduating classes of 1989 and beyond who failed the test in previous years as well as transfer students. The graduating class of 1987 was the first class required to pass the MFMT-II in order to receive a Maryland High School Diploma.

The remainder of this report contains separate sections describing and reporting the development procedures, technical characteristics, and student performance results of the 1988 edition of the MFMT-II. More detailed descriptions of the procedures used in test development, as well as more technical descriptions of psychometric methods used to evaluate functional tests are available from MSDE.

DEVELOPMENT

Declared Competencies Index and the MFMT-II

All Maryland public schools are required to teach the content areas described in the Declared Competencies Index (DCI). The DCI provides the basis for the test specifications and objectives assessed by the MFMT-II. In December 1981, staff from

the Division of Instruction (DOI) and Project Basic produced the Project Basic Instructional Guide, Volume V: Functional Mathematics (MSDE, 1981) designed to help teachers and supervisors develop instructional activities to prepare students for the MFMT-II.

Test Specifications

The Test Specifications (MSDE, 1982) for the MFMT-II define the eligible content for the seven domains of the MFMT-II that include: number concepts, whole number operations, mixed number/fraction operations, decimal operations, measurement, using data, and problem solving. This document describes test specifications that guide item writers in generating new items for both field testing and operational use on subsequent forms of the MFMT-II. It is periodically reviewed and revised according to the general procedures that are available from MSDE. The test specifications committee for the MFMT-II did not meet prior to the development of the 1988 test, as no revisions were proposed or made to the specifications during that time.

Item Development

When necessary, the item writing committee for the MFMT-II convenes to generate new items to assess the objectives in each of the seven domains in the test. Items are checked for readability, item/objective match, and conformance with test specifications. They are reviewed, revised, and field tested prior to inclusion in the MFMT-II item bank. Although no new items were generated by the committee for the 1988 MFMT-II, several existing items from the item bank were revised, reviewed and field tested again on the 1988 MFMT-II. In addition, several items that were revised based on the 1986 test specifications review were field tested in 1987 and used operationally in 1988.

Item Review

In order to be included on an edition of a test in the Maryland Functional Testing Program, items must pass content, bias, psychometric, and editorial reviews. For further details contact MSDE. The operational and field test items included in the 1988 MFMT-II were submitted to these reviews.

Test Forms

Each year, ten forms of the MFMT-II are administered. All ten forms have the same 70 operational items that count toward a student's total score on the test. Each form of the test also includes seven field test items. This format allows the Program Assessment, Evaluation, and Instructional Support Branch (PAEISB) to gather field test data on a total of 70 items each year. For further information on field test procedures and the purposes of field testing, contact MSDE.

TECHNICAL CHARACTERISTICS

Reliability

The reliability of the MFMT-II is evaluated using two statistical indicators: classification consistency and internal consistency coefficients. The MFMT-II is accompanied by a passing score adopted by the Maryland State Board of Education. This score resulted from standard setting procedures that included recommendations from groups of citizens, teachers, administrators, parents, and students from around the state, as well as student field test performance data from the fall of 1984. This passing score corresponds to answering approximately 80 percent of the operational items correctly on a given test and signifies the level of performance required for graduation certification. The standard setting procedures for the MFMT-II are discussed in the MSDE publication entitled Setting a Standard for Standard Setting, (Chafin, 1983). Students whose MFMT-II score is above the passing score are considered to have mastered this set of functional skills, and students scoring below the passing score are considered not to have mastered the content and, therefore, in need of remedial instruction. The reliability or consistency with which the MFMT-II classifies students into these two categories is essential for drawing valid, believable conclusions about a student's mastery of functional mathematics. An indicator of the consistency of classification is an adaptation of a statistic developed by Subkoviak (in Berk, 1984, pp. 278-280), and referred to here as the PG coefficient.

A necessary condition for high classification consistency is the internal consistency of the test. Scores on criterion-referenced tests (CRTs) such as the MFMT-II are used to infer the degree to which students have mastered a domain of knowledge and skill. To make valid inferences from the MFMT-II to the domain of mathematics knowledge and skill defined by the DCI, items must measure only this knowledge and skill. Internal consistency reliability coefficients provide a measure of the extent to which the items on a test are homogeneous, thereby providing one way of describing test reliability. One internal consistency reliability coefficient is the Kuder-Richardson formula number 20, or KR20 (see, for example, Crocker & Algina, 1986; Gronlund, 1982). The reliability and classification consistency of the 1988 MFMT-II are discussed in the following sections.

Classification consistency. Table 1 indicates that the classification consistency of the MFMT-II is excellent: .90 for the students who took the test in 1988. This coefficient reveals that approximately 90 percent of the students who took the test would be classified in the same way (as a passer or failer) on a second concurrent administration of the test. The PG of .90 indicates that about 10 percent of the students who took the test might have been misclassified as passers or failers. It is

likely that students who may have been classified incorrectly are those whose total scale score was just above or just below the passing score of 340.

Table 1

Classification Consistency (PG) and Internal Consistency (KR20) Coefficients

All forms, all grade 9 students (N=47,024)

Domain	No. Items	KR20	PG
1. Number Concepts	10	.73	.83
2. Whole Number Operations	10	.61	.59
3. Mixed Number/Fraction Operations	10	.83	.84
4. Decimal Operations	10	.72	.84
5. Measurement	10	.72	.82
6. Using Data	10	.73	.83
7. Problem Solving	10	.78	.82
Total Test	70	.94	.90

Individual test forms (77 items each form[a])

Form	No. Students	KR20
270	4,587	.86
271	4,575	.87
272	4,495	.86
273	4,573	.87
274	4,552	.86
275	4,504	.86
276	4,611	.86
277	4,541	.86
278	4,545	.86
279	4,555	.86

Note. [a]KR20 coefficients are not available at this time for the 70 operational items.

Internal Consistency. Table 1 contains the internal consistency coefficients of the 1988 MFMT-II forms and items across forms. KR20 coefficients for the individual forms have remained stable and at a high level over the last four years. KR20 coefficients for the 10 forms of the MFMT-II ranged from .86 to .87.

Although each domain of the test has only ten items, KR20 coefficients are satisfactory for all domains, ranging from .61 to .83 for the 1988 test, and have remained fairly stable over the last four years. The KR20 coefficient for the total test is higher than those for the individual domains because domain KR20 coefficients are calculated on fewer items than for the total test. KR20s are lower bound estimates of test reliability.

Validity

The primary aim in developing, administering, scoring, and reporting the MFMT-II is to provide accurate decisions about students' mastery of mathematics skills. The first condition necessary to assure the validity of any test is the demonstration of acceptable levels of classification consistency and reliability as described in the previous section. Evidence of the validity of the MFMT-II includes information about content validity, construct validity, and freedom from bias. Descriptions of the procedures used to help assure that the MFMT-II is administered according to prescribed procedures, methods used to determine content and construct validity, and procedures used to assist in determining that items are free of bias are available from MSDE.

Equivalence of Forms

The issue of equity, or fairness, is of major importance to any test that has a designated passing score and is used to make decisions concerning high school graduation. Students who take the MFMT-II must be assured that the test they are taking is no more easy or difficult to pass than other editions or forms of the test. Thus, equivalence of test forms refers to equivalence of the difficulty of passing a particular form of the test.

Items are selected for inclusion on a given edition of the MFMT-II to help ensure the equivalence of current forms to previous editions. Statistical pre-equating analysis is performed before the test is administered to ensure that the forms are equivalent in difficulty to one another and to previous editions. This analysis also provides a way to ensure comparability of the passing standard. The details of PAEISB's Rasch equating procedure are available from MSDE.

Table 2 shows raw scores (number of items correct) and their corresponding equated scale, domain, and percent correct scores and standard errors. Students who correctly answered 57 items on the 1988 MFMT-II achieved a scale score of 340, which is the lowest possible passing scale score in 1988. This passing raw score is the same as the lowest possible passing raw score of 57 in 1987.

Table 3 contains raw scores, equated scale scores, domain scores, and percent correct scores and standard errors for each domain. To illustrate the use of this table, the panel for domain

Table 2

Lookup Table for the Total Test

Raw Score	Domain Score	Percent Score	Scale Score	Scale Score Standard Error	
0	0	0	182	30	
1	1	1	202	20	
2	3	2	217	15	
3	5	4	226	12	
4	6	5	233	11	
5	8	7	238	10	
6	10	8	243	9	
7	12	10	247	9	
8	13	11	250	8	
9	15	12	254	8	
10	17	14	257	8	
11	18	15	260	7	
12	20	17	262	7	
13	21	18	265	7	
14	23	20	267	7	
15	25	21	269	7	
16	26	22	271	6	
17	28	24	273	6	
18	29	25	275	6	
19	31	27	277	6	
20	32	28	279	6	
21	34	30	281	6	
22	35	31	283	6	
23	37	32	284	6	
24	39	34	286	6	
25	40	35	288	6	
26	41	37	289	6	
27	43	38	291	6	
28	45	40	293	6	
29	45	41	294	6	
30	47	42	296	6	
31	48	44	297	6	
32	50	45	299	6	
33	51	47	300	5	
34	53	48	302	5	
35	54	50	303	5	
36	56	51	305	5	
37	56	52	306	5	
38	59	54	308	5	
39	60	55	309	5	
40	61	57	311	5	
41	62	58	312	6	
42	64	60	314	6	
43	64	61	315	6	
44	67	62	317	6	
45	67	64	318	6	
46	69	65	320	6	
47	70	67	322	6	
48	71	68	323	6	
49	73	70	325	6	
50	75	71	327	6	
51	75	72	328	6	
52	76	74	330	6	
53	78	75	332	6	
54	79	77	334	6	
55	81	78	336	6	
56	81	80	338	7	
Pass Score * 57	** 83	** 81	** 340	* 7	* Pass Score
58	85	82	343	7	
59	85	84	345	7	
60	87	85	348	7	
61	88	87	350	8	
62	89	88	353	8	
63	91	90	357	8	
64	92	91	360	9	
65	93	92	365	10	
66	94	94	370	11	
67	95	95	376	12	
68	97	97	385	15	
69	98	98	400	20	
70	99	100	419	29	

Table 3

Equated Raw, Domain, and Scale Scores and Standard Errors for Each Domain

DOMAIN 1 -NUMBER CONCEPTS

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = .843 STD. DEVIATION = .880

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	6	0	-2.810	1.631	241	33	8.40
1	15	10	-1.628	1.087	264	22	11.44
2	26	20	-.743	.834	282	17	12.84
3	37	30	-.132	.740	294	15	13.30
4	47	40	.380	.696	305	14	13.26
5	55	50	.853	.682	314	14	12.93
6	64	60	1.323	.692	323	14	12.36
7	73	70	1.826	.728	334	15	11.47
8	81	80	2.425	.826	345	17	10.33
9	89	90	3.295	1.081	363	22	8.37
10	96	100	4.464	1.622	386	32	5.25

DOMAIN 2 -WHOLE NUMBERS

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = -1.508 STD. DEVIATION = 1.123

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	2	0	-5.234	1.623	192	32	4.94
1	8	10	-4.064	1.082	216	22	8.75
2	18	20	-3.189	.830	233	17	11.83
3	28	30	-2.581	.738	245	15	13.78
4	39	40	-2.064	.705	256	14	15.07
5	51	50	-1.572	.702	266	14	15.61
6	61	60	-1.065	.726	276	15	15.30
7	72	70	-.499	.782	287	16	13.98
8	83	80	.189	.884	301	18	11.34
9	92	90	1.169	1.132	320	23	7.52
10	97	100	2.451	1.698	346	34	3.84

DOMAIN 3-MIXED NUMBER/FRACTIONS

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = .406 STD. DEVIATION = .499

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	3	0	-3.022	1.599	237	32	5.68
1	10	10	-1.885	1.066	259	21	9.51
2	20	20	-1.046	.805	276	16	12.11
3	29	30	-.481	.701	287	14	13.46
4	39	40	-.016	.662	297	13	14.37
5	48	50	.412	.649	305	13	14.69
6	58	60	.838	.661	314	13	14.55
7	68	70	1.299	.698	323	14	13.83
8	78	80	1.857	.800	334	16	12.62
9	88	90	2.685	1.061	351	21	10.05
10	95	100	3.810	1.591	373	32	6.03

DOMAIN 4 -DECIMALS

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = .426 STD. DEVIATION = 1.221

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	4	0	-3.486	1.658	227	33	7.13
1	13	10	-2.264	1.105	252	22	11.56
2	25	20	-1.337	.862	270	17	13.87
3	36	30	-.676	.772	283	15	14.55
4	48	40	-.106	.739	295	15	14.52
5	59	50	.430	.729	306	15	13.82
6	68	60	.966	.738	316	15	12.61
7	77	70	1.534	.770	328	15	10.96
8	85	80	2.189	.857	341	17	9.04
9	92	90	3.105	1.099	359	22	6.51
10	96	100	4.313	1.649	383	33	3.65

Table 3 (cont.)

Equated Raw, Domain and Scale Scores and Standard Errors for Each Domain

DOMAIN 5 -MEASUREMENT

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = 1.056 STD. DEVIATION = 1.382

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	5	0	-3.187	1.754	233	35	8.67
1	16	10	-1.819	1.169	261	23	12.49
2	30	20	-.765	.918	282	18	13.66
3	42	30	-.024	.811	297	16	13.44
4	52	40	.582	.751	309	15	12.74
5	61	50	1.126	.728	320	15	11.95
6	69	60	1.655	.732	330	15	11.09
7	77	70	2.213	.767	341	15	10.07
8	84	80	2.867	.860	354	17	8.79
9	92	90	3.797	1.110	373	22	6.72
10	97	100	5.028	1.664	398	33	3.89

DOMAIN 6 -USING DATA

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = .130 STD. DEVIATION = 1.044

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	5	0	-3.566	1.618	226	32	7.76
1	13	10	-2.403	1.079	249	22	10.72
2	24	20	-1.534	.829	266	17	11.84
3	33	30	-.926	.740	278	15	12.38
4	43	40	-.403	.710	289	14	12.81
5	51	50	.095	.707	299	14	13.08
6	61	60	.604	.724	309	14	13.06
7	71	70	1.158	.767	320	15	12.58
8	80	80	1.806	.855	333	17	11.26
9	90	90	2.722	1.100	351	22	8.68
10	96	100	3.930	1.649	376	33	5.02

DOMAIN 7 -PROBLEM SOLVING

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = .573 STD. DEVIATION = .727

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	4	0	-2.937	1.603	238	32	7.27
1	12	10	-1.795	1.068	261	21	11.06
2	24	20	-.950	.812	278	16	12.81
3	34	30	-.373	.713	290	14	13.42
4	43	40	.109	.677	299	14	13.74
5	52	50	.559	.668	308	13	13.68
6	62	60	1.012	.682	317	14	13.30
7	70	70	1.505	.727	327	15	12.58
8	80	80	2.099	.823	339	16	11.26
9	89	90	2.966	1.079	356	22	8.94
10	96	100	4.131	1.619	380	32	5.44

1 indicates that a student who correctly answered 8 of 10 items, equivalent to a scale score of 345, can be considered to have mastered the domain. However, pass-fail status is not determined by performance within an individual domain: a student may perform poorly on a given domain, yet pass the test by performing well in other domains. One reason for determining pass-fail status only at the total test level is that longer tests are generally more reliable than shorter tests. This relationship between test length and test reliability is reflected in the scale score standard errors (SEss) for domain one. A scale score of 345, based on ten items in the domain, has a standard error of 17--considerably higher than the standard error of 7 for a scale score of 345 based on 70 items (see Table 2).

Item "Bias" Analysis

For the MFTP, test item "bias" is defined as differential item functioning (DIF) in contrasting groups of examinees. Potential DIF for reference and focal groups -- males versus females and whites versus blacks -- is sought in two different stages. First, during the item development stage, items are subjected to judgmental reviews by a bias committee. Any item judged to potentially function differently for one group than for another is deleted from the item pool or revised before field testing. Second, all operational and field test items are statistically analyzed for potential DIF after each census test administration.

The statistical approach used to detect DIF is the transformed item difficulties (TID) method described by Osterlind (1985, p. 34). The purpose of the analysis is to identify items that may be functioning differently for reference versus focal groups. If an item is found to be more difficult for one group than for another, the item is flagged as possibly "biased." Flagged items are put on probationary status and their content and format are reviewed at the next meeting of the bias committee for possible causes of DIF. Flagged items can be used on subsequent editions of the MFMT-II, but if they are flagged statistically on a subsequent edition as possibly "biased" they are removed from the item bank. If a flagged item is not flagged in its next two administrations, it is removed from probation.

The TID approach studies the interaction of two or more groups' performance with items. TID uses a relative standard: that is, an item is considered to be biased when it is comparatively more difficult for one group to answer correctly than it is for the other. Item p-values are calculated for each of two groups that have been matched on ability. These p-values are converted to z-values and then transformed to delta values with a mean of 13 and a standard deviation of 4. The delta values for each pair of comparison groups for each item are plotted. Potential DIF is sought by (1) examining 99 percent confidence intervals constructed around a regression line for each plot, (2) examining standardized DIF values, and (3) inspecting plots visu-

ally. "DIF" is defined as the distance of each plotted item from the major axis line. Unacceptable "DIF" is defined in the MFTP as any item with a DIF value outside of the 99 percent confidence interval, or a standardized DIF value greater than plus or minus two (Angoff, 1982, p. 98; Rudner, 1977). Plots provide descriptive information that aids in interpretation of DIF results.

DIF for males versus females. Table 4 and Figure 1 contain the 1988 MFMT-II DIF information for males (reference group) versus females (focal group). DIF and standardized DIF values can be found in the two right hand columns in Table 4 under the heading "DIF*" and "STD. DIF." Any item with a DIF value outside of the 99 percent confidence interval or a standardized DIF value greater than plus or minus two may be functioning differently for one group than it is for another and is flagged for further study. DIF values between -.75 and .75 are within 99 percent confidence limits for the male-female comparisons. Non-significant differences between values are likely due to random fluctuations in subgroup performance and measurement error.

Only one DIF value in Table 4 falls outside of the 99 percent confidence interval; this item also has a standardized DIF value greater than 2.0. Operational sequence number 46, (actual item number 50) has a DIF value of -.89 and a standardized DIF of -3.06. In addition, operational sequence number 45 (actual item number 49) has a standardized DIF value of -2.38 although its DIF value of -.69 is within acceptable limits. These two items have been put on probationary status and will be discussed in future reports. The DIF criterion has resulted in identification of two items which are slightly more difficult for females than they are for males, but still fall within the item cluster on the plot. Figure 1 shows the plot of the standardized delta values; however no points on the plot are outside of the cluster. Finally, the correlation between the two groups' DIF values for all items is .98, indicating a strong degree of linear relationship.

DIF for whites versus blacks. Table 5 and Figure 2 contain the DIF information for whites (reference group) versus blacks (focal group). For this comparison, the 99 percent confidence interval limits are -.80 and .80. Operational sequence number 45 (actual item number 49) has a DIF value of -1.01 and a standardized DIF of -3.27, both of which are outside of acceptable ranges. This item appears to be easier for whites than it is for blacks. In addition, operational sequence numbers 29 and 46 (actual item numbers 31 and 50, respectively) have DIF values within the acceptable range but standardized DIF values of 2.20 and -2.46, respectively. Item 29 appears to be easier for blacks; item 46 easier for whites. These items have been put on probation and will be discussed in future reports. All of the remaining DIF and standardized DIF values in Table 5 fall within acceptable ranges, indicating that no additional item on the 1988 MFMT-II shows statistical evidence of potential race "bias." Figure 2 shows that all points on the plot are within the major

Table 4

DIF and Standardized DIF Values for Bias Analysis: Males versus Females

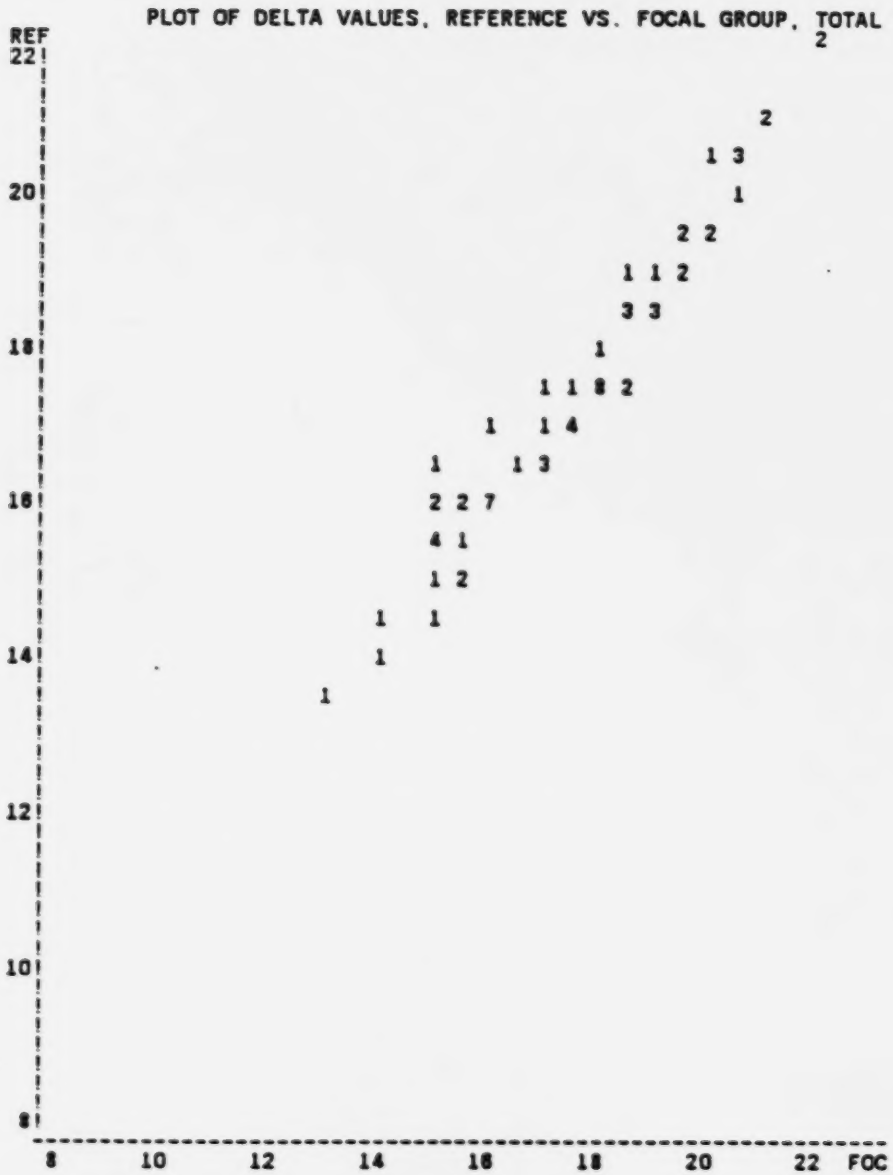
SEQUENCE NUMBER	P-VALUE REF	P-VALUE FOC	Z-VALUE REF	Z-VALUE FOC	DELTA REF	DELTA FOC	DIF*	STD. DIF
I001	.95	.96	1.65	1.75	19.60	20.00	.13	.45
I002	.96	.97	1.75	1.88	20.00	20.52	.19	.67
I003	.99	.99	2.33	2.33	22.32	22.32	-.28	-.95
I004	.99	.99	2.33	2.33	22.32	22.32	-.28	-.95
I005	.97	.96	1.88	1.75	20.52	20.00	-.54	-1.87
I006	.98	.98	2.05	2.05	21.20	21.20	-.22	-.76
I007	.95	.95	1.65	1.65	19.60	19.60	-.14	-.49
I008	.94	.95	1.55	1.65	19.20	19.60	.15	.52
I009	.82	.84	.92	.98	16.88	16.96	.20	.67
I010	.85	.87	1.04	1.13	17.16	17.52	.23	.78
I011	.92	.91	1.41	1.34	18.64	18.36	-.28	-.98
I012	.84	.86	.98	1.08	16.96	17.32	.24	.81
I013	.72	.71	.58	.55	15.32	15.20	-.01	-.03
I014	.78	.76	.77	.71	16.08	15.84	-.13	-.44
I015	.78	.76	.77	.71	16.08	15.84	-.13	-.44
I016	.82	.82	.92	.92	16.88	16.88	.00	.02
I017	.92	.93	1.41	1.48	18.64	18.92	.10	.34
I018	.89	.89	.50	.50	15.00	15.00	.09	.31
I019	.77	.76	.74	.71	15.96	15.84	-.04	-.14
I020	.83	.84	.95	.98	16.80	16.96	.11	.37
I021	.92	.94	1.41	1.55	18.64	19.20	.29	1.00
I022	.97	.97	1.88	1.88	20.52	20.52	-.18	-.64
I023	.88	.89	1.08	1.23	17.32	17.92	.38	1.32
I024	.88	.90	1.17	1.28	17.68	18.12	.25	.88
I025	.94	.95	1.55	1.65	19.20	19.60	.15	.52
I026	.91	.92	1.34	1.41	18.36	18.64	.11	.39
I027	.82	.84	.92	.98	16.88	16.96	.20	.67
I028	.71	.75	.55	.67	15.20	15.68	.41	1.40
I029	.78	.78	.77	.77	15.84	15.84	.21	.73
I030	.71	.72	.55	.58	15.20	15.32	.18	.55
I031	.87	.88	1.13	1.23	17.52	17.92	.24	.81
I032	.87	.90	1.13	1.28	17.52	18.12	.37	1.28
I033	.88	.91	1.17	1.34	17.68	18.36	.42	1.44
I034	.84	.87	.98	1.13	16.96	17.52	.37	1.28
I035	.78	.77	.77	.74	15.84	15.96	.13	.44
I036	.88	.90	1.17	1.28	17.68	18.12	.25	.88
I037	.85	.88	1.04	1.17	17.16	17.68	.34	1.16
I038	.88	.88	1.08	1.08	17.32	17.32	-.03	-.09
I039	.89	.89	1.23	1.23	17.92	17.92	-.08	-.20
I040	.73	.70	.61	.52	15.44	15.08	-.18	-.62
I041	.88	.85	1.17	1.04	17.68	17.16	-.40	-1.38
I042	.98	.98	2.05	2.05	21.20	21.20	-.22	-.76
I043	.77	.71	.74	.55	15.96	15.20	-.48	-1.64
I044	.72	.74	.58	.64	15.32	15.56	-.24	.81
I045	.84	.77	.98	.74	16.96	15.96	-.89	-2.38
I046	.81	.71	.88	.55	16.52	15.20	-.89	-3.08
I047	.62	.58	.30	.20	14.20	13.80	-.14	-.50
I048	.58	.52	.15	.05	13.80	13.20	-.11	-.39
I049	.64	.58	.36	.20	14.44	13.80	-.32	-1.10
I050	.78	.71	.77	.55	16.08	15.20	-.57	-1.95
I051	.95	.96	1.65	1.75	19.60	20.00	.13	.45
I052	.95	.95	1.65	1.65	19.60	19.60	-.14	-.49
I053	.97	.97	1.88	1.88	20.52	20.52	-.19	-.64
I054	.93	.92	1.48	1.41	18.92	18.64	-.30	-1.03
I055	.78	.74	.77	.64	16.08	15.56	-.32	-1.10
I056	.97	.97	1.88	1.88	20.52	20.52	-.19	-.64
I057	.94	.94	1.55	1.55	19.20	19.20	-.12	-.42
I058	.74	.68	.64	.47	15.56	14.88	-.40	-1.39
I059	.77	.75	.74	.67	15.96	15.68	-.15	-.52
I060	.79	.79	.81	.81	16.24	16.24	.03	.09
I061	.81	.83	.88	.95	16.52	16.80	.20	.70
I062	.66	.70	.41	.52	14.84	15.08	.41	1.40
I063	.78	.78	.77	.77	16.08	16.08	.03	.12
I064	.92	.92	1.41	1.41	18.64	18.64	-.09	-.32
I065	.91	.94	1.34	1.55	18.36	19.20	.48	1.70
I066	.88	.90	1.17	1.28	17.68	18.12	.25	.88
I067	.88	.91	1.17	1.34	17.68	18.36	.42	1.44
I068	.73	.70	.61	.52	15.44	15.08	-.18	-.62
I069	.87	.89	1.13	1.23	17.52	17.92	.24	.81
I070	.87	.89	1.13	1.23	17.52	17.92	.24	.81

TOTAL RECORDS-REFERENCE = 21338
TOTAL RECORDS-----FOCAL = 21379

STANDARD DEV.-REFERENCE = 1.94
STANDARD DEV.-----FOCAL = 2.08

VARIANCE------REFERENCE = 3.77
VARIANCE----------FOCAL = 4.33

MEAN DELTA----REFERENCE = 17.50
MEAN DELTA--------FOCAL = 17.55



SLOPE = .93
INTERCEPT = 1.14
CORRELATION COEFFICIENT= .98

MEAN DIF = -.00
STANDARD DEVIATION DIF= .29

99% C.I. = -.00 + OR - 2.58 * .29 = (-.75, .75)

*DIF = DISTANCE FROM THE REGRESSION LINE

Figure 1. Plot of standardized delta values for males versus females.

Table 5

DIF and Standardized DIF Values for Bias Analysis: Whites versus Blacks

SEQUENCE NUMBER	P-VALUE REF	P-VALUE FOC	Z-VALUE REF	Z-VALUE FOC	DELTA REF	DELTA FOC	DIF*	STD. DIF
I001	.94	.95	1.55	1.65	19.20	19.60	.18	.57
I002	.95	.96	1.65	1.75	19.60	20.00	.16	.53
I003	.98	.99	2.05	2.33	21.20	22.32	.61	1.98
I004	.98	.98	2.05	2.05	21.20	21.20	-.16	-.53
I005	.95	.95	1.65	1.65	19.60	19.60	-.11	-.37
I006	.97	.98	1.88	2.05	20.52	21.20	.33	1.08
I007	.94	.94	1.55	1.55	19.20	19.20	-.10	-.32
I008	.92	.92	1.41	1.41	18.64	18.64	-.08	-.27
I009	.78	.78	.81	.77	16.24	16.08	-.11	-.37
I010	.83	.81	.95	.88	16.80	16.52	-.22	-.70
I011	.88	.88	1.23	1.17	17.92	17.68	-.22	-.73
I012	.81	.84	.88	.99	16.52	16.96	.28	.94
I013	.82	.88	.30	.47	14.20	14.88	.53	1.72
I014	.89	.73	.50	.81	15.00	15.44	.34	1.10
I015	.70	.74	.52	.64	15.08	15.58	.38	1.18
I016	.76	.78	.71	.77	15.64	16.08	.17	.57
I017	.90	.90	1.28	1.28	18.12	18.12	-.08	-.21
I018	.82	.88	.30	.47	14.20	14.88	.53	1.72
I019	.71	.73	.55	.81	15.20	15.44	.20	.63
I020	.78	.83	.81	.95	16.24	16.80	.38	1.24
I021	.91	.92	1.34	1.41	18.36	18.64	.12	.39
I022	.96	.96	1.75	1.75	20.00	20.00	-.13	-.41
I023	.83	.85	.95	1.04	16.80	17.16	.23	.73
I024	.85	.87	1.04	1.13	17.16	17.52	.21	.70
I025	.93	.93	1.48	1.48	18.92	18.92	-.09	-.29
I026	.89	.89	1.23	1.23	17.92	17.92	-.06	-.19
I027	.80	.77	.84	.74	16.36	15.98	-.28	-.92
I028	.63	.70	.33	.52	14.32	15.08	.58	1.89
I029	.67	.75	.44	.67	14.76	15.68	.68	2.20
I030	.82	.85	.30	.39	14.20	14.58	.31	1.01
I031	.82	.83	.92	.95	16.68	16.80	.06	.21
I032	.82	.84	.92	.99	16.68	16.96	.17	.57
I033	.84	.85	.99	1.04	16.96	17.16	.11	.36
I034	.79	.79	.81	.81	16.24	16.24	-.00	-.01
I035	.88	.89	.47	.50	14.88	15.00	.12	.40
I036	.85	.86	1.04	1.08	17.16	17.32	.08	.25
I037	.80	.82	.84	.92	16.36	16.68	.21	.69
I038	.80	.82	.84	.92	16.36	16.68	.21	.69
I039	.84	.86	.99	1.08	16.96	17.32	.22	.72
I040	.84	.83	.36	.33	14.44	14.32	-.03	-.09
I041	.85	.83	1.04	.95	17.16	16.80	-.28	-.92
I042	.98	.98	2.05	2.05	21.20	21.20	-.16	-.53
I043	.88	.86	.47	.41	14.88	14.64	-.13	-.41
I044	.87	.88	.44	.47	14.76	14.88	.13	.41
I045	.81	.70	.88	.52	16.52	15.08	-1.01	-3.27
I046	.75	.85	.67	.39	15.68	14.56	-.76	-2.48
I047	.52	.54	.05	.10	13.20	13.40	.23	.75
I048	.48	.41	-.05	-.23	12.80	12.08	-.39	-1.27
I049	.51	.48	.03	-.05	13.12	12.80	-.12	-.40
I050	.70	.83	.52	.33	15.08	14.32	-.49	-1.60
I051	.94	.94	1.55	1.55	19.20	19.20	-.10	-.32
I052	.94	.93	1.55	1.48	19.20	18.92	-.28	-.95
I053	.96	.96	1.75	1.75	20.00	20.00	-.13	-.41
I054	.92	.89	1.41	1.23	18.64	17.92	-.58	-1.88
I055	.71	.68	.55	.47	15.20	14.88	-.19	-.62
I056	.95	.95	1.65	1.65	19.60	19.60	-.11	-.37
I057	.93	.91	1.48	1.34	18.92	18.36	-.48	-1.55
I058	.65	.61	.39	.28	14.56	14.12	-.25	-.82
I059	.70	.68	.52	.47	15.08	14.88	-.10	-.34
I060	.72	.71	.58	.55	15.32	15.20	-.06	-.19
I061	.78	.77	.77	.74	16.08	15.98	-.08	-.27
I062	.60	.58	.25	.20	14.00	13.80	-.07	-.23
I063	.73	.69	.61	.50	15.44	15.00	-.28	-.92
I064	.89	.88	1.23	1.17	17.92	17.68	-.22	-.73
I065	.88	.90	1.17	1.28	17.68	18.12	.25	.82
I066	.85	.85	1.04	1.04	17.16	17.18	-.03	-.11
I067	.84	.85	.99	1.04	16.96	17.18	.11	.36
I068	.66	.60	.41	.25	14.64	14.00	-.39	-1.28
I069	.83	.84	.95	.99	16.80	16.96	.09	.29
I070	.82	.84	.92	.99	16.68	16.96	.17	.57

TOTAL RECORDS-REFERENCE = 9706
TOTAL RECORDS-----FOCAL = 10014

STANDARD DEV.-REFERENCE = 2.06
STANDARD DEV.-----FOCAL = 2.16

VARIANCE------REFERENCE = 4.25
VARIANCE----------FOCAL = 4.65

MEAN DELTA----REFERENCE = 18.79
MEAN DELTA--------FOCAL = 18.83

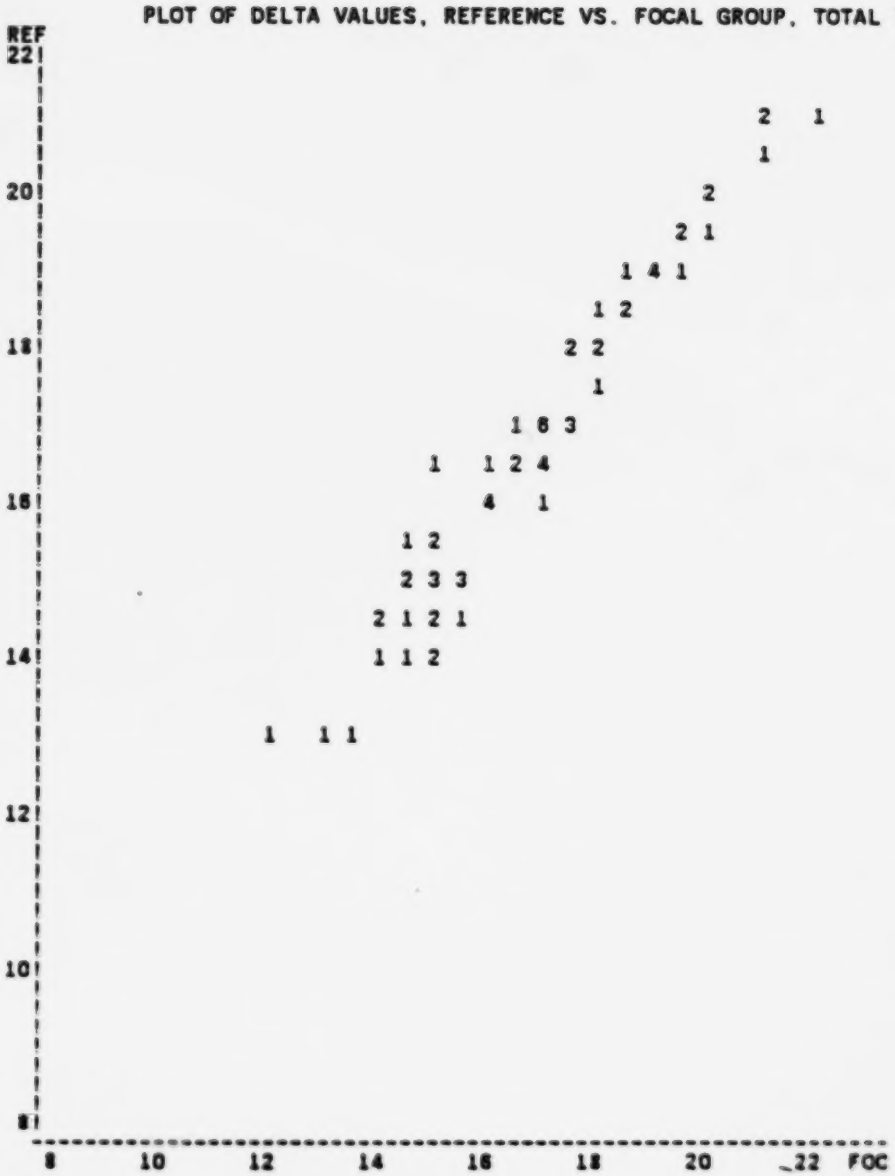


Figure 2. Plot of standardized delta values for whites versus blacks.

cluster. Finally, the correlation between the two groups' DIF values for all items is .98, indicating a strong degree of linear relationship.

Analysis of Item 48

It was discovered during Fall testing that item 48 (operational sequence number 44) does not fully conform to item specifications for "area items" on the MFMT-II. (On area items, students are required to calculate the area of a polygon.) Specifications require that the interior of the polygon for area items be shaded, while the interior of the polygon for perimeter items not be shaded. The polygon that appears with item 48 on all forms of the 1988 MFMT-II is not shaded; also, the polygon for this item was not shaded when it was field tested in 1986. Since this item does not fully conform to item specifications for the MFMT-II, a question arises as to whether it is appropriate to include performance on the item in student scores on the Fall 1988 MFMT-II. Determination of whether to include or exclude this item from scoring, however, must be considered in conjunction with considerations for content validity and reliability of scores.

In order to answer this question prior to scoring the 1988 MFMT-II, item statistics from item 48 were compared with those for all other area items in the MFMT-II bank. The goal of this comparison was to determine if item difficulty was unduly affected by omitting the interior shading in the polygon. Adjusted Rasch item difficulty values are used in this analysis because of their invariance property. The bank item difficulty value for item 48, based on field testing in 1986, is 1.602; the bank value from its operational administration in 1988 is 1.428. In comparison, bank item difficulty values for the other ten area items in the MFMT-II bank range from a low of 0.740 to a high of 3.559. That is, the difficulty values for all other area items vary widely, while the difficulty of item 48 is somewhere in the middle of this range. It appears that the difficulty of item 48 is not unreasonable or unduly affected by its lack of full conformity to item specifications. Based on these observations item 48 was included in scoring the 1988 MFMT-II. This item will not appear in future editions of the MFMT-II until it conforms with MFMT-II test specifications and has been field tested again.

RESULTS

Pass-rates. The percentage of students who took the MFMT-II for the first time and passed decreased slightly from 1987 (see Table 6 below). The pass-rates for the past seven years for first-time takers as well as retested students are also presented.

Table 6

Pass-Rates for the MFMT-II: 1982-1988

	Percent Passing						
	School Year						
Grade	1982-83	1983-84	1984-85	1985-86	1986-87	1987-88	1988-89
9	39.6[a]	60.8	61.0	64.9	65.5	68.3	67.4
10	NA	NA	33.6	41.8	40.2	44.4	39.4
11	NA	NA	NA	45.8	48.4	51.5	48.5
12	NA	NA	NA	NA	56.4	60.4	56.4

Note. Percentages include fall administration only.
NA=Not available. [a]Indicates a "no-fault" administration in which test performance does not affect graduation from high school, and for which no retesting occurs.

Response Proportions. Table 7 contains the response proportions and p-values for each operational item on the 1988 MFMT-II. The response proportion indicates the proportion of students selecting each possible response for each item on the MFMT-II. Response proportions are used to evaluate the quality of the incorrect responses (or distractors) and can be useful in identifying confusing or miskeyed items. The p-value indicates the percentage of students who answered an item correctly. The p-values for the 1988 MFMT-II range from .53 (operational sequence number 48, actual item number 52) to .99 (operational sequence and actual item number 3). Experience has defined the acceptable range of p-values as .30 to .70 for norm-referenced tests. However, this convention applies less strictly to criterion-referenced tests that may include items outside of this range in order to ensure adequate coverage of relevant content areas. In particular, it is appropriate for basic skills tests like the MFMT-II to contain a number of "easier" items -- with p-values greater than .7 -- to ensure accurate measurement of low-achieving students and meaningful diagnostic domain scores.

Summary Results. The mean raw, domain, and scale scores and their standard deviations for the 1988 MFMT-II are presented in Table 8. The mean scale score of 352.9 represents a .1 increase over 1987's mean scale score of 352.8. Finally, Table 9 presents the distribution of frequencies of each score point for the 1988 MFMT-II.

Table 7

Percentages of Students Who Answered Each Item Correctly (P-Values) and Response Proportions

SEQUENCE NUMBER	RESPONSE PROPORTIONS A	B	C	D	OTHER	P-VALUE
I001	.01	.03	.95	.00	.00	.95
I002	.96	.01	.02	.01	.00	.96
I003	.99	.00	.00	.01	.00	.99
I004	.01	.01	.98	.00	.00	.98
I005	.02	.01	.01	.96	.00	.96
I006	.01	.98	.00	.01	.00	.98
I007	.95	.03	.02	.01	.00	.95
I008	.02	.93	.02	.02	.00	.93
I009	.13	.03	.01	.82	.00	.82
I010	.01	.10	.85	.03	.00	.85
I011	.91	.06	.00	.03	.00	.91
I012	.09	.04	.02	.84	.00	.84
I013	.16	.09	.04	.70	.00	.70
I014	.76	.05	.12	.07	.00	.76
I015	.03	.76	.12	.08	.00	.76
I016	.15	.02	.02	.81	.00	.81
I017	.92	.03	.02	.04	.00	.92
I018	.04	.17	.68	.11	.00	.68
I019	.11	.05	.09	.75	.00	.75
I020	.83	.10	.04	.03	.00	.83
I021	.92	.02	.02	.04	.00	.92
I022	.96	.02	.01	.01	.00	.96
I023	.07	.02	.05	.86	.00	.86
I024	.01	.88	.06	.04	.00	.88
I025	.03	.01	.94	.02	.00	.94
I026	.06	.03	.90	.01	.00	.90
I027	.05	.82	.11	.02	.00	.82
I028	.09	.03	.17	.71	.00	.71
I029	.13	.05	.76	.06	.00	.76
I030	.70	.09	.16	.05	.00	.70
I031	.87	.05	.07	.02	.00	.87
I032	.03	.04	.07	.87	.00	.87
I033	.03	.88	.07	.02	.00	.88
I034	.06	.84	.07	.03	.00	.84
I035	.06	.12	.07	.75	.00	.75
I036	.03	.08	.88	.01	.00	.88
I037	.07	.85	.04	.04	.00	.85
I038	.02	.09	.85	.03	.00	.85
I039	.88	.04	.06	.01	.00	.88
I040	.22	.71	.04	.03	.00	.71
I041	.01	.09	.03	.86	.00	.86
I042	.01	.98	.01	.01	.00	.98
I043	.10	.15	.02	.73	.00	.73
I044	.02	.71	.21	.06	.00	.71
I045	.07	.05	.80	.09	.00	.80
I046	.75	.08	.14	.03	.00	.75
I047	.16	.09	.59	.16	.00	.59
I048	.53	.04	.22	.21	.00	.53
I049	.24	.11	.05	.59	.00	.59
I050	.73	.24	.02	.01	.00	.73
I051	.03	.95	.01	.01	.00	.95
I052	.94	.01	.02	.03	.00	.94
I053	.01	.00	.03	.96	.00	.96
I054	.01	.04	.92	.03	.00	.92
I055	.03	.75	.11	.10	.00	.75
I056	.96	.01	.01	.01	.00	.96
I057	.04	.01	.93	.01	.00	.93
I058	.21	.09	.69	.00	.00	.69
I059	.18	.06	.74	.01	.00	.74
I060	.02	.15	.06	.77	.00	.77
I061	.03	.81	.09	.07	.00	.81
I062	.17	.67	.11	.04	.01	.67
I063	.77	.10	.09	.04	.00	.77
I064	.03	.91	.04	.02	.00	.91
I065	.02	.03	.91	.03	.00	.91
I066	.03	.88	.04	.05	.00	.88
I067	.06	.02	.03	.88	.00	.88
I068	.07	.12	.70	.10	.01	.70
I069	.02	.87	.08	.03	.00	.87
I070	.05	.05	.03	.87	.00	.87

TOTAL RECORDS = 47024

Table 8

Mean Raw, Domain, and Scale Scores and Standard Deviations (in parentheses): Grade 9 (N=47,024)

Domain	Maximum Raw Score	Raw Score	Domain Score	Scale Score
Number Concepts	10	8.0 (2.1)	79.8 (17.8)	352.6 (30.0)
Whole Number Operations	10	9.4 (1.1)	92.5 (9.6)	332.4 (21.2)
Mixed Number/ Fraction Operations	10	8.4 (2.3)	80.4 (21.7)	348.0 (30.2)
Decimal Operations	10	8.5 (1.9)	85.9 (14.8)	355.3 (29.5)
Measurement	10	7.3 (2.3)	77.7 (18.0)	350.7 (32.9)
Using Data	10	8.6 (1.8)	84.8 (15.9)	350.3 (28.1)
Problem Solving	10	8.3 (2.1)	81.3 (19.2)	350.5 (29.6)
All Domains	70	58.3 (10.7)	-- --	352.9 (27.6)

Table 9

Frequency Distribution of Total Scores

RAW SCORE	SCALE SCORE	PERCENT SCORE	FREQUENCY	FREQUENCY PERCENTAGE	CUMULATIVE FREQUENCY	CUMULATIVE PERCENTAGE
0	182	.0	0	.0	0	.0
1	202	1.4	1	.0	1	.0
2	217	2.9	0	.0	1	.0
3	226	4.3	1	.0	2	.0
4	233	5.7	0	.0	2	.0
5	238	7.1	0	.0	2	.0
6	243	8.6	1	.0	3	.0
7	247	10.0	4	.0	7	.0
8	250	11.4	0	.0	7	.0
9	254	12.9	4	.0	11	.0
10	257	14.3	0	.0	11	.0
11	260	15.7	3	.0	14	.0
12	262	17.1	9	.0	23	.0
13	265	18.6	9	.0	32	.1
14	267	20.0	9	.0	41	.1
15	269	21.4	14	.0	55	.1
16	271	22.9	23	.0	78	.2
17	273	24.3	32	.1	110	.2
18	275	25.7	51	.1	161	.3
19	277	27.1	49	.1	210	.4
20	279	28.6	51	.1	261	.6
21	281	30.0	68	.1	329	.7
22	283	31.4	82	.2	411	.9
23	284	32.9	79	.2	490	1.0
24	286	34.3	111	.2	601	1.3
25	288	35.7	127	.3	728	1.5
26	289	37.1	112	.2	840	1.8
27	291	38.6	123	.3	963	2.0
28	293	40.0	142	.3	1105	2.3
29	294	41.4	122	.3	1227	2.6
30	296	42.9	163	.3	1390	3.0
31	297	44.3	157	.3	1547	3.3
32	299	45.7	179	.4	1726	3.7
33	300	47.1	185	.4	1911	4.1
34	302	48.6	240	.5	2151	4.6
35	303	50.0	219	.5	2370	5.0
36	305	51.4	200	.4	2570	5.5
37	306	52.9	236	.5	2806	6.0
38	308	54.3	251	.5	3057	6.5
39	309	55.7	303	.6	3360	7.1
40	311	57.1	326	.7	3686	7.8
41	312	58.6	378	.8	4064	8.6
42	314	60.0	393	.8	4457	9.5
43	315	61.4	412	.9	4869	10.4
44	317	62.9	450	1.0	5319	11.3
45	318	64.3	477	1.0	5796	12.3
46	320	65.7	514	1.1	6310	13.4
47	322	67.1	570	1.2	6880	14.6
48	323	68.6	642	1.4	7522	16.0
49	325	70.0	715	1.5	8237	17.5
50	327	71.4	787	1.7	9024	19.2
51	328	72.9	799	1.7	9823	20.9
52	330	74.3	885	1.9	10708	22.8
53	332	75.7	932	2.0	11640	24.8
54	334	77.1	1140	2.4	12780	27.2
55	336	78.6	1218	2.6	13998	29.8
56	338	80.0	1281	2.7	15279	32.5
57	340	81.4	1381	2.9	16660	35.4
58	343	82.9	1539	3.3	18199	38.7
59	345	84.3	1587	3.4	19786	42.1
60	348	85.7	1851	3.9	21637	46.0
61	350	87.1	1966	4.2	23603	50.2
62	353	88.6	2056	4.4	25659	54.6
63	357	90.0	2305	4.9	27964	59.5
64	360	91.4	2437	5.2	30401	64.6
65	365	92.9	2680	5.7	33081	70.3
66	370	94.3	2804	6.0	35885	76.3
67	376	95.7	2989	6.4	38874	82.7
68	385	97.1	3031	6.4	41905	89.1
69	400	98.6	2907	6.2	44812	95.3
70	419	100.0	2212	4.7	47024	100.0

References

Angoff, W. H. (1982). Use of difficulty and discrimination indices for detecting item bias. In R. A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore, MD: Johns Hopkins University Press.

Chafin, A. E. (1983). Setting a standard for standard setting. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. NY: CBS College Publishing.

Gronlund, N. E. (1982). Constructing achievement tests (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Maryland State Department of Education (1980). Declared competencies index. Baltimore, MD: Author.

Maryland State Department of Education. (1981). Project basic instructional guide, Volume V: Functional mathematics. Baltimore, MD: Author.

Maryland State Department of Education. (1982). Test specifications: Maryland functional mathematics tests. Baltimore, MD: Author.

Osterlind, S. J. (1985). Test item bias. Beverly Hills: Sage Publications.

Subkoviak, M. (1984). Estimating the reliability of mastery-nonmastery classifications. In R. A. Berk (Ed.), A guide to criterion-referenced test construction. Baltimore: The Johns Hopkins University Press.

Rudner, L. M. (1977). An evaluation of select approaches for biased item modification. Unpublished doctoral dissertation, Catholic University of America, Washington, D.C.

LWB28000